

ANLY482 Analytics Practicum

Meeting Minutes

Date:	17/02/2015
Time:	1530h
Venue:	SIS Meeting Room 4.1
Attendees:	Prof KAM Tin Seong CHENG Fu Mei LEONG Wai Sum Lynette SEOW Hui Xin
Absent with Apologies:	–
Minutes Taker:	LEONG Wai Sum

Agenda:

Topic	Exploratory Data Analysis
Content	<ul style="list-style-type: none">• Missing data analysis<ul style="list-style-type: none">○ E.g. <i>"inbound_contract_code"</i>○ Why are there missing data/so many missing values?○ How many percent constitutes a significantly high amount of missing data?○ Need to change workflow to make sure the database is properly filled up○ Steps should be taken to reduce number of missing values○ Recommendations: need to fill up gaps for future analytics use○ Inform the sponsor/management of the individual missing values○ Transformation and weigh it• Summary statistics for locations fields<ul style="list-style-type: none">○ Can report on e.g. top 5, 10 of <i>"Map Dest Srva"</i>○ Depends on what is being analysed○ <i>"Map Origin Srva"</i> vs <i>"Local Revenue"</i>○ Use bar chart/histogram○ Use <i>"Sum"</i> function instead of data points○ Customers who have a lot of variance in transactions, some months with very big variance and some months small○ Get total revenue, and get e.g. top 10 customers (monetary)○ Find consistent customers (similar revenue every month), the proportion of customers whose purchases are straight lines throughout the month/week○ The customers whose purchases fluctuate (very high/low in certain months/weeks), variation in purchase○ Look at yearly/monthly/weekly○ Use <i>"Sum"</i> instead of <i>"Median"</i> in order to observe patterns and see the difference, for e.g. <i>"Local Revenue"</i> and <i>"Billed Weight"</i>○ Change field to months○ Subsequently use box-plot to show variations and distributions of months,

	<p>see if it is the same or different throughout the month; which customers have higher variation and which are the more stable accounts</p> <ul style="list-style-type: none"> • Customers <ul style="list-style-type: none"> ○ Focus should be on how to identify the different groups of customers ○ Target customers at the end of the day ○ Look at distribution of each customer (account number) ○ For each customer, calculate mean and standard deviation (SD); those with high SD are those that fluctuate (week/month/year) • Sales Channel <ul style="list-style-type: none"> ○ Can one customer use more than 1 sales channel? ○ If so, what proportion uses 1, 2 or 3 sales channel(s)? ○ Break down the customers using the number of sales channels they use ○ For each sales channel, how do they perform overtime? E.g. Sales channel A is used by 50% of the customers ○ However, average revenue per customer may not be that way; should also see in terms of proportion ○ Compare absolute total revenue, average revenue per transaction, and also revenue per customer (total revenue over number of customers to get an average) ○ Understand segments better: number of sales channel(s) used, proportion of the different numbers of sales channels used ○ Monetary → Pareto (80:20) rule for revenue vs account number ○ (Run with account managers in company is dangerous and risky as they may bring away customers along with them if they join competitor companies) ○ How many %, aggregate out by customers, to segregate customers who consistently give high/low revenue ○ Show statistics for revenue, accompanied with distribution, e.g. 25th, 50th, 75th percentiles ○ Show in bar graph, not line ○ Put lines to show where is 25th, 50th, 75th percentiles ○ Look at total, not mean ○ Look at total first, then later compare mean and median to get an idea of the distribution (e.g. skewness) ○ Check accounts with 0 revenue but yet having transactions (Summary > Frequency – N Rows) <ul style="list-style-type: none"> ▪ To single out ▪ May be due to bundle sales: package items together for discounts (may be e.g. promotions, free trial), utilise 2-3 times then do not use the service again already – to check
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Topic	Cluster Analysis
Content	<ul style="list-style-type: none"> • RFM model <ul style="list-style-type: none"> ○ RFM binning

	<ul style="list-style-type: none"> ○ Start with the conventional model of RFM as the variables ○ Derive variables → RFM index (individually from transaction data) ○ Come up with indicators (derived variables) to represent R, F and M <ul style="list-style-type: none"> ▪ Based on variables, how can recency (R) be found? E.g. 1 - 12 (based on months) ▪ Most recent can be this month, week, or even day ▪ If based on days, most recent is 1, latest will be 365 (or 366 in a leap year); for no purchase at all throughout the year, it should not be 0, need to put 999 or something bigger than 365/366 (e.g. 380, 400) as no purchase should logically be assigned the biggest number ▪ As for frequency (F), can take row count from transaction data ▪ Monetary (M) is traditionally looking at the total sum; so M is the total revenue per customer (“Account”) ○ Weakness of scoring method of RFM model presented in the readings: subjective approach, so not to follow ○ Instead, use clustering (analytics technique) to do customer segmentation: put R, F, M into the clustering process and examine the results ○ Need to transform and apply standardisation (e.g. normalisation) ○ However, the consistency of contribution may be an important factor as well ○ Should we use RFM like what other businesses are doing? <ul style="list-style-type: none"> ▪ It does not tell about the customers with fluctuating purchases ▪ See variance, may be able to add in additional insights ▪ Will it be a better model, more refined, better targeting? Or will it just make the model more confusing without adding much value? ● Other factors/variables besides RFM <ul style="list-style-type: none"> ○ Insights can be derived from EDA ○ High variance in transactions ○ Consistency in orders
--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Follow-up Actions:

S/N	Task	Person Assigned	Due Date
1	EDA: <i>Customer</i> field	Fu Mei	21/02/2015, 1159h
2	EDA: <i>Channel, Product, Industry</i> fields	Wai Sum	21/02/2015, 1159h
3	EDA: <i>Locations (Owner, Origin, Destination)</i> fields	Lynette	21/02/2015, 1159h